

DDI Moving Forward

The uptake of DDI Lifecycle since its publication in 2008-2009 has been quite dramatic. The lifecycle concept has resonated with the community, and we have seen several new audiences learn about and adopt DDI.

We have also heard a lot about ways in which the specification can be improved and enhanced, and these ideas mesh with our goals in these four high-level areas:

- Complete data life cycle coverage
- Broadened focus for new research domains
- Robust and persistent data model (for the metadata), extension possibilities, implementation for different technical domains
- Simpler specification that is easier to understand and use including better documentation

After publishing DDI Lifecycle 3.2 in the fall of 2012, the time seems right to start to address these improvements. The Technical Implementation Committee seeks approval to begin this important work. Further, we propose a workshop to be held at Schloss Dagstuhl, Wadern, Germany, October 22-26, 2012, to focus on gathering requirements for and modeling this new DDI version.

Specific Changes Envisioned

It is critical that DDI Lifecycle be based on a UML data model as the canonical version. This model can then be expressed in XML Schema, RDF/OWL Ontology, relational database schema, and other languages. Such an abstract data model will make it easier to interact with other disciplines and other standards, to understand the specification, to develop and maintain it in a consistent and structured way, and to enable software development that is less dependent on specific DDI versions.

In addition, we envision that as it moves forward DDI-L will address:

- Abstraction of data capture/collection/source. The current data collection module is questionnaire-centric. We should also be able to describe register data and data in the natural and health sciences (i.e., from technical devices or from laboratory analyses). There would be an abstract layer for data collection with the possibilities for “plug-ins” to handle different types of data.
- New content on sampling, survey implementation, weighting, and paradata coming out of the Survey Design and Implementation Group
- New content developed by the Qualitative Working Group
- Framework for data and metadata quality
- Framework for access to data and metadata
- Process (work flow) description across the data life cycle, including support for automation and replication

- Integration with existing standards like GSBPM/GSIM, SDMX, CDISC, Triple-S
- Disclosure review and remediation
- Data management planning
- Development of standard queries and/or interface specifications (such as REST) which are needed to allow for interoperable services based on the DDI standard information model

DDI Codebook

We also anticipate the creation of a similar model for DDI Codebook, perhaps as a subset of the DDI-L model. This would facilitate the work of those organizations/developers who are working with DDI-C or with both development lines of DDI. There are several organizations that are taking up DDI-C either as a primary structure or as an input to a DDI-L system.

Why Now?

After the experience of developing and using the initial DDI-L structure, we find that continued improvement of the development line is limited by the lack of a data model. Further, we are experiencing pressure for changes from several directions at once. We need to integrate new content from substantive working groups, which requires new approaches to the design of the specification. We are also being asked to interact with standards such as SDMX. Right now DDI is seen as a strong and viable standard for describing and sharing data within the National Statistical Framework on an international level, and we would like it to remain so. Our responsiveness to our current and potential user base as well as DDI's openness to member involvement is seen as a positive feature of DDI as a standard. Further, we are seeing huge changes in the larger "ecosystem" for data, many of which were outlined in our 2010 meeting in Ithaca. DDI needs to adapt to meet some of these challenges.

Frequently Asked Questions

Is this a third development line for DDI?

No. This is a continuation of the DDI-Lifecycle development line. The objectives of the structural change are to improve its functionality from a technical perspective; provide a model that is more accessible to users (i.e., data professionals and developers), particularly those outside of the traditional social sciences; and create a more flexible base for further development and coverage extensions.

What do these changes mean for those currently taking up DDI Lifecycle?

Users choosing to take up DDI Lifecycle prior to these new developments have the same choice as any user of a specific DDI version. If the chosen version meets the needs for a given project, people should continue to use it. If the functionality of the newer DDI Lifecycle version is needed, then using the newer specification will make more sense. Users migrating existing DDI instances will be provided with information detailing the changes that need to be made to their content to be valid in the new version.

Will all work stop on DDI Codebook? What if there is a bug?

No. Bugs will be corrected. New functionality or coverage expansion are not anticipated.

Is the anticipated structural change on the same scale as DDI Version 2.1 to Version 3.0?

No. The move from Codebook to Lifecycle was a major conceptual shift. Codebook was designed for a particular usage and environment. Within that environment it works well as evidenced by the new wave of uptake for this standard. Lifecycle sought to encompass a broader range of uses within a more complex and wide-ranging environment. Version 3.0 was our first attempt to meet this new requirement. Versions 3.1 and 3.2 have been exercises in fine tuning this structure as more developers start working with it. The structural work for Lifecycle being discussed here reflects what has been learned through the first four years of user experience in terms of what is needed to make it easier to implement and integrate within a broad range of systems, uses, and organizations.

What is the timeframe?

This is not an overnight process -- it will take a minimum of two years. Prior to starting on this work TIC will migrate to a new system for tracking their work, including improved access to bug tracking, current discussions, and work activities, and improved access for feedback and review during the development process. Developers who need to anticipate what changes will be made will have the ability to access information throughout the development process rather than waiting for the end product.